

Trabajo Práctico Final 2019

Joaquín Manuel

Ejercicio 1

Enunciado:

En cada ronda de un juego, un jugador gana \$1, con probabilidad p , o pierde \$1, con probabilidad $1 - p$. El jugador comienza con $\$k$. El juego se detiene cuando el jugador pierde todo su dinero o gana un total de $\$n$, donde $n > k$. Las fortunas sucesivas del jugador forman una cadena de Markov en $0, 1, \dots, n$ con $X_0 = k$.

- (a) Simular la ruina del jugador para una inversión inicial de \$2, jugando un juego justo ($p = (1 - p)$).
- (b) Estimar la probabilidad de que el jugador llegue a la ruina antes de ganar \$5.
- (c) Construir la matriz de transición para la cadena de markov asociada. Estimar la probabilidad en a) calculando potencias altas de la matriz.
- (d) Comparar los resultados con la probabilidad exacta.

Resolución:

- (a) Un juego justo es en el cual el participante tiene la misma probabilidad de ganar \$1 como de perder \$1 por lo tanto $p = 1 - p = 0.5$

Código:

```
1 gamble <- function(k,n,p) {  
2   while (k > 0 & k < n)  
3     k <- k + sample(c(-1,1),1,prob=c(1-p,p))  
4     k == 0  
5 }
```

Resultados:

```
1 > mean(replicate(1000, gamble(2, 5, 1/2)))
2 [1] 0.61
3 > mean(replicate(1000, gamble(2, 10, 1/2)))
4 [1] 0.802
5 > mean(replicate(1000, gamble(2, 50, 1/2)))
6 [1] 0.958
7 > mean(replicate(1000, gamble(2, 500, 1/2)))
8 [1] 0.998
9 > mean(replicate(100, gamble(250, 500, 1/2)))
10 [1] 0.5
```

Conclusión:

Se puede observar que a medida que aumenta n la probabilidad de que el jugador llegue a la ruina es mayor, esto se debe a que la inversión inicial es de \$2 que está a menor distancia de 0 que de n . Si $k = \frac{n}{2}$ la probabilidad sería la misma de ganar que de perder.

- (b) La probabilidad de llegar a la ruina antes de ganar \$5 para cualquier $n \leq 5$ es la misma que si $n = 5$ ya que si $n = 5$ el estado 5 será absorbente y una vez que se llega a él no se podrá llegar a la ruina. Por lo tanto aquí podemos crear una cadena de markov donde $E = \{0, 1, 2, 3, 4, 5\}$ que cumpla con lo pedido donde la matriz de transición para esta cadena será:

$$q = 1 - p$$
$$P = \begin{matrix} & \begin{matrix} 0 & 1 & 2 & 3 & 4 & 5 \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix} & \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ q & 0 & p & 0 & 0 & 0 \\ 0 & q & 0 & p & 0 & 0 \\ 0 & 0 & q & 0 & p & 0 \\ 0 & 0 & 0 & q & 0 & p \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{matrix}$$

Como $p = 1 - p = 0.5$, tenemos que :

$$P = \begin{matrix} & \begin{matrix} 0 & 1 & 2 & 3 & 4 & 5 \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix} & \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{matrix}$$

Código:

```
1 r = rep(0,1000)
2 for(i in 1:1000){
3   x = rmarkovchain(mc,n=100,t0="2")
4   for(j in 1:100){
5     if(x[j] == "0"){
6       r[i] = 1
7       break
8     }
9     if(x[j] == "5") break
10  }
11 }
12 mean(r)
13 mean(replicate(1000, gamble(2, 5, 1/2)))
```

Resultados:

```
1 > mean(r)
2 [1] 0.61
3 > mean(replicate(1000, gamble(2, 5, 1/2)))
4 [1] 0.61
```

Conclusión:

Luego de simular 1000 trayectorias siendo la inversión inicial \$2, y sacando el promedio podemos estimar que la probabilidad de que el jugador llegue a la ruina antes de ganar \$5 es 0.61.

- (c) Tomando la segunda simulación del item (a) donde $n = 10$ y $k = 2$ La matriz de transición es:

$$P = \begin{matrix} & \begin{matrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \end{matrix} & \begin{pmatrix} 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.5 & 0.0 & 0.5 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.5 & 0.0 & 0.5 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.5 & 0.0 & 0.5 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.5 & 0.0 & 0.5 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.5 & 0.0 & 0.5 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.5 & 0.0 & 0.5 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.5 & 0.0 & 0.5 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.5 & 0.0 & 0.5 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.5 & 0.0 & 0.5 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \end{pmatrix} \end{matrix}$$

Código:

```
1 transitionProbability(mc^100,"2","0")
```

Resultados:

```
1 [1] 0.7976061
```

- (d) La probabilidad exacta para $p = 1/2$ es $\frac{n-k}{n} = \frac{10-2}{10} = \frac{8}{10} = 0.8$. Podemos observar que ambos procesos dieron un resultado muy cercano al teórico (0.802 con la rutina y 0.797 calculando potencias altas).

Utilizando la teoría de cadenas de Markov también es posible llegar al resultado teórico calculando $F(2,0)$ que es la probabilidad de partiendo de \$2 llegar a la ruina en un número finito de jugadas.

Para eso pasamos P a forma canónica, obtenemos Q , B y calculamos:

$$S = (I - Q)^{-1}$$

Luego obtenemos

$$G = S * B$$

	0	10
1	0.9	0.1
2	0.8	0.2
3	0.7	0.3
4	0.6	0.4
5	0.5	0.5
6	0.4	0.6
7	0.3	0.7
8	0.2	0.8
9	0.1	0.9

Donde $G(2,0) = F(2,0) = 0.8$

Ejercicio 2

Enunciado:

En aplicaciones de seguridad informática, un **honeypot** es una herramienta dispuesta en una red o sistema informático para ser el objetivo de un posible ataque informático, y así poder detectarlo y obtener información del mismo y del atacante. Los datos del **honeypot** son estudiados utilizando cadenas de Markov. Se obtienen datos desde una base de datos central y se observan ataques contra cuatro puertos {80, 135, 139, 445} durante un año. Los estados

de la cadena son los cuatro puertos y se incluye un nodo indicando que ningún puerto está siendo atacado (NA). Los datos se monitorean semanalmente y el puerto más atacado durante la semana es guardado. La matriz de transición para la cadena estimada para los ataques semanales es la siguiente:

$$E = \{80, 135, 139, 145, NA\}$$

$$P = \begin{matrix} & \begin{matrix} 80 & 135 & 139 & 145 & NA \end{matrix} \\ \begin{matrix} 80 \\ 135 \\ 139 \\ 145 \\ NA \end{matrix} & \begin{pmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 8/13 & 3/13 & 1/13 & 1/13 \\ 1/16 & 3/16 & 3/8 & 1/4 & 1/8 \\ 0 & 1/11 & 4/11 & 5/11 & 1/11 \\ 0 & 1/8 & 1/2 & 1/8 & 1/4 \end{pmatrix} \end{matrix}$$

La distribución inicial es $a = (0, 0, 0, 0, 1)$

- Después de dos semanas, ¿Cuáles son los puertos con más y menos probabilidad de ser atacados?
- Encontrar la distribución límite (si es que existe) de los puertos atacados. Justificar.

Resolución:

- Para saber la distribución de probabilidad luego de 2 semanas se calcula el producto entre la distribución inicial y la matriz de transición a 2 pasos es decir $a.P^2$.

Código:

```
1 a <- c(0,0,0,0,1)
2 semana2 = a * (mc^2) # mc es la cadena de markov
3 semana2
```

Resultados:

```
1      80      135      139      145      NA
2 [1,] 0.03125 0.2132867 0.3868007 0.2226836 0.145979
```

Conclusión:

Se puede observar que el puerto con mayor probabilidad de ser atacado es el **139** y el puerto con menor probabilidad de ser atacados es el **80**.

- (b) La cadena de Markov es irreducible por no tener ningún subconjunto propio cerrado de estados y es aperiódica ya que $P(X_{n+1} = 139|X_n = 139) > 0$, por lo tanto para obtener su distribución límite basta con calcular su distribución estacionaria.

Código:

```
1 is.irreducible(mc)
2 period(mc)
3 steadyStates(mc)
```

Resultados:

```
1 [1] TRUE
2 [1] 1
3           80           135           139           145           NA
4 [1,] 0.02146667 0.2669333 0.3434667 0.2273333 0.1408
```

Ejercicio 3

Enunciado:

Dans et al. (2012) estudian el comportamiento de delfines en presencia de embarcaciones turísticas en la Patagonia Argentina. Para ello postulan un modelo de cadena de Markov, con espacio de estados conformado por las 5 actividades primarias de los delfines: socializar (s), viajar (t), merodear (m), alimentarse (f), descansar (r), obteniendo la siguiente matriz de transición:

$$E = \{s, t, m, a, r\}$$

$$P = \begin{matrix} & \begin{matrix} s & t & m & a & r \end{matrix} \\ \begin{matrix} s \\ t \\ m \\ a \\ r \end{matrix} & \begin{pmatrix} 0.84 & 0.11 & 0.01 & 0.04 & 0 \\ 0.03 & 0.80 & 0.04 & 0.10 & 0.03 \\ 0.01 & 0.15 & 0.70 & 0.07 & 0.07 \\ 0.03 & 0.19 & 0.02 & 0.75 & 0.01 \\ 0.03 & 0.09 & 0.05 & 0 & 0.83 \end{pmatrix} \end{matrix}$$

- Clasificar los estados.
- Estimar la distribución a largo plazo de la actividad de los delfines.
- Simular y graficar una trayectoria de dicho proceso.

Resolución:

- (a) La cadena de Markov es irreducible, aperiódica y tiene una cantidad de estados finitos por lo tanto $C_s = C_t = C_m = C_a = C_r = E$, entonces todos sus estados son recurrentes y no hay absorbentes ya que si no tendría un subconjunto cerrado propio.

Código:

```
1 is.irreducible(mc)
2 period(mc)
```

Resultados:

```
1 [1] TRUE
2 [1] 1
```

- (b) La cadena de Markov es irreducible por no tener ningún subconjunto propio cerrado de estados y es aperiódica ya que $P(X_{n+1} = m | X_n = m) > 0$, por lo tanto para obtener su distribución límite basta con calcular su distribución estacionaria.

Código:

```
1 steadyStates(mc)
```

Resultados:

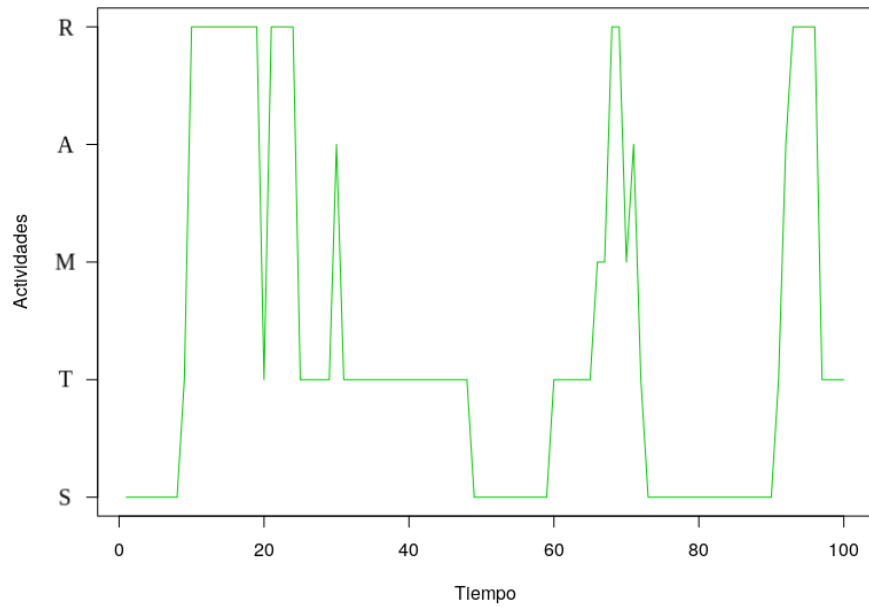
```
1           s           t           m           a           r
2 [1,] 0.1478358 0.4149254 0.0955597 0.2163806 0.1252985
```

Conclusión:

La distribución a largo plazo de la actividad de los delfines es:

$$\pi = (0.1478358, 0.4149254, 0.0955597, 0.2163806, 0.1252985)$$

(c) La gráfica representa una simulación de 100 pasos:

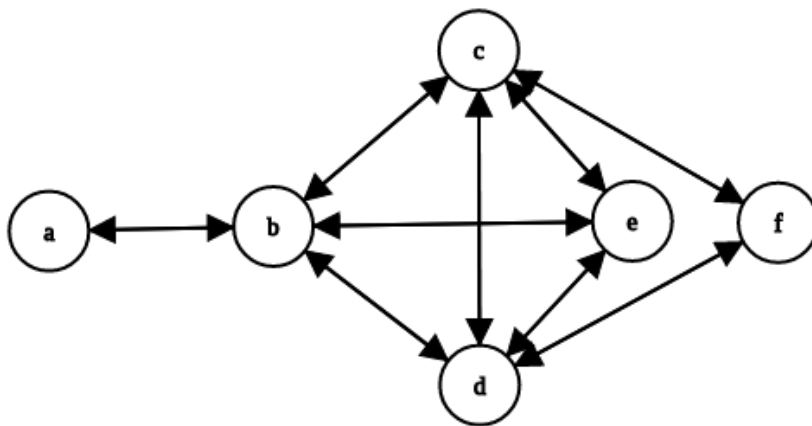


Se puede ver que no hay demasiado cambio de actividades, esto se debe a que la probabilidad de que los delfines sigan en el mismo estado en el que estaban es bastante alta en todos los estados (probabilidades en la diagonal de la matriz P).

Ejercicio 4

Enunciado:

Simular 50 pasos de caminata aleatoria en el grafo correspondiente a la Figura.



- (a) Repetir la simulación 10 veces. ¿Cuántas terminaron en el vértice c ?
- (b) Comparar con el resultado exacto de la probabilidad a largo plazo de visitar a c .

Resolución:

$$E = \{a, b, c, d, e, f\}$$
$$P = \begin{matrix} & \begin{matrix} a & b & c & d & e & f \end{matrix} \\ \begin{matrix} a \\ b \\ c \\ d \\ e \\ f \end{matrix} & \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1/4 & 0 & 1/4 & 1/4 & 1/4 & 0 \\ 0 & 1/4 & 0 & 1/4 & 1/4 & 1/4 \\ 0 & 1/4 & 1/4 & 0 & 1/4 & 1/4 \\ 0 & 1/3 & 1/3 & 1/3 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 & 0 & 0 \end{pmatrix} \end{matrix}$$

- (a) **Código:**

```
1 cont = 0
2 for(i in 1:10){
3     x = rmarkovchain(mc,n=50)
4     if(tail(x, n = 1) == "c") cont = cont + 1
5 }
6 cont
```

Resultados:

1 [1] 2

Conclusión:

Al repetir la simulación de caminata aleatoria 10 veces, en 2 de esas simulaciones se terminó en el vértice c.

- (b) La cadena de Markov es irreducible y es aperiodica por lo tanto para obtener su distribución límite basta con calcular su distribución estacionaria. La distribución límite es:

$$\pi = (0.05555556, 0.2222222, 0.2222222, 0.2222222, 0.1666667, 0.1111111)$$

Esta distribución límite nos da la probabilidad a largo plazo de visitar c, la cual es 0.2222222, si la aproximamos por nuestro experimento obtendríamos que es $2/10 = 0.2$, se logra un resultado cercano al teórico pero no exacto.

Ejercicio 5

Enunciado:

Hay k canciones en el reproductor de música de María. El reproductor está seteado en modo shuffle, en el cual las canciones se eligen aleatoriamente, de forma uniforme, con reemplazo. Por lo tanto, las repeticiones de canciones es posible. Sea X_n el número de canciones no repetidas que han sido escuchadas después de la n -ésima reproducción.

- (a) Mostrar que $\{X_n : n \in N\}$ es una cadena de Markov y determinar la matriz de transición.
- (b) Si María tiene cuatro canciones en su reproductor de música, encontrar la probabilidad de que todas las canciones sean escuchadas después de 6 reproducciones.

Resolución:

- (a) Una cadena de Markov es un proceso estocástico con conjuntos de instantes de observación, T , infinito numerable y espacio de estados, E , discreto. Que además cumple:

$$\forall i_0, i_1, \dots, i_{n-1}, i, j$$

$$P(X_{n+1} = j | X_0 = i_0, \dots, X_{n-1} = i_{n-1}, X_n = i) = P(X_{n+1} = j | X_n = i)$$

La familia de variables aleatorias definidas en el enunciado son un proceso estocástico donde T es la cantidad de reproducciones por lo tanto $T = N$ y el espacio $E = 0, 1, 2, \dots, k$ son los naturales (con el 0) hasta k ya que no se pueden escuchar mas de k canciones distintas.

Dado X_n podemos definir X_{n+1} como $X_n + Y_{n+1}$ donde $Y_{n+1} \sim Be(\frac{k-X_n}{k})$.

$$Y_{n+1} = \begin{cases} 1 & \text{Con probabilidad } p = \frac{k-X_n}{k} \\ 0 & \text{Con probabilidad } 1 - p \end{cases}$$

Se puede observar que X_{n+1} sólo depende del estado anterior X_n por lo tanto:

$$\begin{aligned} P(X_{n+1} = j | X_0 = i_0, \dots, X_{n-1} = i_{n-1}, X_n = i) &= \\ P(X_n + Y_{n+1} = j | X_0 = i_0, \dots, X_{n-1} = i_{n-1}, X_n = i) &= \\ P(Y_{n+1} = j - i | X_0 = i_0, \dots, X_{n-1} = i_{n-1}, X_n = i) &= \\ P(Y_{n+1} = j - i | X_n = i) &= \\ P(X_{n+1} = j | X_n = i) & \end{aligned}$$

$\{X_n : n \in N\}$ es una cadena de Markov cuya matriz de transición será:

$$P = \begin{matrix} & \begin{matrix} 0 & 1 & 2 & 3 & \dots & k \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \\ 3 \\ \vdots \\ k \end{matrix} & \begin{pmatrix} 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & \frac{1}{k} & \frac{k-1}{k} & 0 & \dots & 0 \\ 0 & 0 & \frac{2}{k} & \frac{k-2}{k} & \dots & 0 \\ 0 & 0 & 0 & \frac{3}{k} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & \frac{k}{k} \end{pmatrix} \end{matrix}$$

(b) Para armar la matriz de transición necesitamos analizar cada uno de los estados posibles:

- El estado 0 marca que no se escucho ninguna canción, entonces al realizar una reproducción si o si habrá escuchado una canción que no se escuchó antes por lo tanto pasa al estado 1 con probabilidad 1.
- En los estados 1, 2, 3, es posible pasar al estado siguiente si escucha una canción "nueva" o quedarse en el mismo si escucha una que escucho antes.
- El estado 4 será absorbente ya que solo tiene 4 canciones distintas y una vez que las escuchó ya no puede "olvidarse" de haberlas escuchado.

En el estado i donde $1 \leq i \leq 3$ la probabilidad de escuchar una nueva canción y pasar al estado siguiente es $\frac{4-i}{4}$ que es la cantidad de formas de tomar una "nueva" sobre el total de formas de tomar una canción cualquiera. La probabilidad de quedarse en el mismo estado (escuchar una canción que ya escucho antes) es $\frac{i}{4}$.

$$P = \begin{matrix} & \begin{matrix} 0 & 1 & 2 & 3 & 4 \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{matrix} & \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & \frac{1}{4} & \frac{3}{4} & 0 & 0 \\ 0 & 0 & \frac{2}{4} & \frac{2}{4} & 0 \\ 0 & 0 & 0 & \frac{3}{4} & \frac{1}{4} \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{matrix}$$

Asumo que la distribución inicial es $a = (1, 0, 0, 0, 0)$ en la que no ha escuchado ninguna canción.

Para obtener la probabilidad de luego de 6 reproducciones haber escuchado todas las canciones debemos calcular $F_6(0, 4) + F_5(0, 4) + F_4(0, 4)$ es decir la suma de las probabilidades de llegar por primera vez en 4, 5 y 6 pasos ya que es imposible llegar en menos de 4 pasos. Pero al ser el estado 4 absorbente toda trayectoria que llegó al estado 4 en 4 o 5 pasos también llegará al estado 4 en 6 pasos por lo tanto $F_6(0, 4) + F_5(0, 4) + F_4(0, 4) = P^6(0, 4)$ y luego:

$$P(X_6 = 4 | X_0 = 0) = P^6(0, 4) = 0.3808594$$

Calculamos $F_6(0, 4) + F_5(0, 4) + F_4(0, 4)$ y simulamos para verificar:

Código:

```

1 P <- function(n,i,j){
2   transitionProbability(mc^n,i,j)
3 }
4
5 F <- function(n,i,j,E){
6   if(n==1) return (P(1,i,j))
7   else {
8     sum = 0
9     for(b in 1:length(E)){
10      if(E[b] != j) sum = sum + P(1,i,E[b]) *
11                          F(n-1,E[b],j,E)
12    }
13    return (sum)
14  }
15 }
16

```

```

17 r = rep(0,1000)
18 for(i in 1:1000){
19   x = rmarkovchain(mc,n=6,t0="0")
20   for(j in 1:6){
21     if(x[j] == "4"){
22       r[i] = 1
23       break
24     }
25   }
26 }

```

Resultados:

```

1 > F(6,"0","4",states) +
2 F(5,"0","4",states) + F(4,"0","4",states)
3 [1] 0.3808594
4 > mean(r)
5 [1] 0.384

```

Ejercicio 6

Enunciado:

Se tiran 5 dados y se ponen a un lado aquellos que mostraron un 6. Los restantes se lanzan nuevamente y se repite el procedimiento, poniendo a un lado aquellos dados que muestran un 6, y así sucesivamente. Sea X_n el número de dados en los que salió 6 después de n tiradas.

- Describir la matriz de transición P para la cadena de Markov.
- Encontrar la probabilidad de obtener todos 6 en tres jugadas.
- ¿Cómo se espera que sea P^{100} ? Confirmar la respuesta utilizando R .

Resolución:

- Los estados serán $E = \{0, 1, 2, 3, 4, 5\}$ ya que son las cantidades posibles de dados con 6. Además desde cada estado i se podrá llegar a un estado j con una probabilidad mayor a 0 solo si $i \leq j$.

$$P(X_{n+1} = j | X_n = i) > 0 \iff i \leq j$$

$$P_{ij} = 0 \quad \forall i > j$$

Defino una variable aleatoria Y : "Cantidad de dados que mostraron un 6 al tirar n dados." . $Y \sim Bi(n, p)$ donde p es la probabilidad de que un dado muestre un 6 por lo tanto $p = 1/6$.

$$P(Y = y) = \binom{n}{y} p^y (1 - p)^{n-y} = \binom{n}{y} \frac{1}{6}^y \frac{5}{6}^{n-y}$$

Con esta variable aleatoria se puede calcular la probabilidad de pasar de un estado a otro en la cadena de Markov, por ejemplo si partimos del estado 1 entonces $n = 4$ ya que se volverán a tirar 4 dados, si queremos saber $P(1, i)$ con $1 \leq i$ calculamos $P(Y = i - 1)$

Código:

```
1 Biny <- function(n,p,y)
2   choose(n,y) * p^y * (1-p)^(n-y)
3
4 P = matrix( rep( 0, len=36), nrow = 6)
5
6 for(i in 1:6)
7   for(j in 1:6)
8     if(i<=j) P[i,j] = Biny(5-i+1,1/6,j-i)
9
10 P
```

Resultados:

$$P = \begin{bmatrix} 0.40187 & 0.40187 & 0.16075 & 0.03215 & 0.00321 & 0.00012 \\ 0.00000 & 0.48225 & 0.38580 & 0.11574 & 0.01543 & 0.00077 \\ 0.00000 & 0.00000 & 0.57870 & 0.34722 & 0.06944 & 0.00462 \\ 0.00000 & 0.00000 & 0.00000 & 0.69444 & 0.27777 & 0.02777 \\ 0.00000 & 0.00000 & 0.00000 & 0.00000 & 0.83333 & 0.16666 \\ 0.00000 & 0.00000 & 0.00000 & 0.00000 & 0.00000 & 1.00000 \end{bmatrix}$$

- (b) Aquí ocurre lo mismo que en el ejercicio anterior en el item (b), el estado 5 es absorbente por lo tanto $F_3(0, 5) + F_2(0, 5) + F_1(0, 5) = P^3(0, 5)$

$$P(X_3 = 5 | X_0 = 0) = P^3(0, 5) = 0.01327206$$

La simulación da como resultado 0.014.

- (c) Se espera que:

$$P^{100} = \begin{matrix} & \begin{matrix} 0 & 1 & 2 & 3 & 4 & 5 \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix} & \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{matrix}$$

Ya que tiene solo un estado absorbente que es el 5, y el resto son transitorios por lo tanto a largo plazo se espera que se termine en el único conjunto cerrado que tiene la cadena, entonces la distribución límite se esperaría que sea $\pi = (0, 0, 0, 0, 0, 1)$ y como consecuencia la matriz anterior.

Resultado de R:

$$P^{100} = \begin{bmatrix} 2.5e-40 & 1.0e-31 & 1.7e-23 & 1.4e-15 & 6.0e-08 & 0.999 \\ 0.0e+00 & 2.1e-32 & 7.0e-24 & 8.7e-16 & 4.8e-08 & 1.000 \\ 0.0e+00 & 0.0e+00 & 1.7e-24 & 4.3e-16 & 3.6e-08 & 1.000 \\ 0.0e+00 & 0.0e+00 & 0.0e+00 & 1.4e-16 & 2.4e-08 & 1.000 \\ 0.0e+00 & 0.0e+00 & 0.0e+00 & 0.0e+00 & 1.2e-08 & 1.000 \\ 0.0e+00 & 0.0e+00 & 0.0e+00 & 0.0e+00 & 0.0e+00 & 1.000 \end{bmatrix}$$

Conclusión:

Podemos ver que se obtiene algo muy parecido a lo que se esperaba, ya que la última columna tiene todos unos (a excepción de 1 que tiene un número muy cercano a 1) y las demás celdas de la matriz tienen números muy pequeños cercanos a 0.

Ejercicio 7

Enunciado:

Considerar la caminata aleatoria $\{0, \dots, k\}$, la cual se mueve a la izquierda o a la derecha con probabilidades p y q respectivamente. Si el proceso está en 0, transiciona a 1 en el siguiente paso. Si el proceso está en k , transiciona a $k-1$ en el siguiente paso. Esto se llama *caminata aleatoria con límites reflectantes*. Asumir que $k=3$, $q=1/4$, $p=3/4$ y la distribución inicial es uniforme.

- Calcular la matriz de transición.
- Encontrar $P(X_7 = 1 | X_0 = 3, X_2 = 2, X_4 = 2)$
- Encontrar $P(X_3 = 1, X_5 = 3)$

Resolución:

- La matriz de transición es:

$$P = \begin{matrix} & \begin{matrix} 0 & 1 & 2 & 3 \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \\ 3 \end{matrix} & \begin{pmatrix} 0 & 1 & 0 & 0 \\ p=3/4 & 0 & q=1/4 & 0 \\ 0 & p=3/4 & 0 & q=1/4 \\ 0 & 0 & 1 & 0 \end{pmatrix} \end{matrix}$$

- (b) Como $P(X_1 = 2|X_0 = 3) = 1$ sabemos que si $X_0 = 3$ entonces $X_1 = 2$ pero al ser una cadena de Markov con período 2 sabemos que:

$$P(X_2 = 2|X_1 = 2) = 0$$

con lo cual $P(X_0 = 3, X_2 = 2, X_4 = 2) = 0$, pero la probabilidad condicionada está definida como:

$$P(A|B) = \frac{P(A, B)}{P(B)} \text{ con } P(B) > 0$$

Por lo tanto no tiene sentido esa expresión ya que no tiene sentido condicionar con un suceso imposible.

Si modificamos el ejercicio a:

$$P(X_7 = 1|X_1 = 3, X_2 = 2, X_4 = 2)$$

Por la definición de cadena de Markov:

$$P(X_7 = 1|X_1 = 3, X_2 = 2, X_4 = 2) = P(X_7 = 1|X_4 = 2) = P^3(2, 1)$$

$$P^3(2, 1) = 0.890625$$

(c)

$$\begin{aligned} P(X_3 = 1, X_5 = 3) &= P(X_5 = 3|X_3 = 1)P(X_3 = 1) = \\ &= P^2(1, 3) (aP^3)(1) = 0.02856445 \end{aligned}$$

Donde $a = (1/4, 1/4, 1/4, 1/4)$ por ser una distribución inicial uniforme.

Ejercicio 8

Enunciado:

Los administradores de datos de una universidad desarrollaron un modelo markoviano para simular los índices de graduación en la institución. Los estudiantes pueden abandonar, repetir un año o pasar al año siguiente. Todos tienen un 3% de chance de repetir de año. Aquellos que se encuentran en primer o segundo año, tienen una chance del 6% de abandonar. Para estudiantes de tercer y cuarto año, el índice de abandono es de 4%.

- (a) Clasificar los estados.
- (b) Establecer la matriz de transición de un paso.
- (c) Determinar el número promedio de años que un estudiante que ingresa en primer año permanece en la institución antes de abandonar o recibirse.

Resolución:

- (a) Los estados son $E = \{A, 1, 2, 3, 4, R\}$ donde A es Abandonar la carrera y R es recibirse. Los estados A y R son estados absorbentes, y los demás estados $\{1, 2, 3, 4\}$ son transitorios ya que no pertenecen a ningún conjunto cerrado irreducible.
- (b) La matriz de transición de un paso es:

$$P = \begin{matrix} & \begin{matrix} A & 1 & 2 & 3 & 4 & R \end{matrix} \\ \begin{matrix} A \\ 1 \\ 2 \\ 3 \\ 4 \\ R \end{matrix} & \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 6/100 & 3/100 & 91/100 & 0 & 0 & 0 \\ 6/100 & 0 & 3/100 & 91/100 & 0 & 0 \\ 4/100 & 0 & 0 & 3/100 & 93/100 & 0 \\ 4/100 & 0 & 0 & 0 & 3/100 & 93/100 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{pmatrix}$$

- (c) Código:

```
1 sum = 0
2 l = 10000
3 for(i in 1:l){
4   x = rmarkovchain(mc,n=20,t0=1)
5   for(j in 1:length(x)){
6     if(x[j] == "A" || x[j] == "R"){
7       sum = sum + j - 1
8       break
9     }
10  }
11 }
12 l + sum/l
13
14 Q = matrix(c( 3/100 , 91/100 , 0 , 0 ,
15              0 , 3/100 , 91/100 , 0 ,
16              0 , 0 , 3/100 , 93/100 ,
17              0 , 0 , 0 , 3/100 )
18           ,byrow=TRUE,nrow=4)
19
20 N = ginv(diag(4)-Q)
21 sum(N[1,])
```

Resultados:

```
1 [1] 3.7816
2 [1] 3.7753
```

Conclusión:

Luego de realizar la simulación 10000 veces se obtiene que el numero promedio de años que un estudiante permanece en la institución es 3.78 años. El resultado es muy cercano al resultado teórico que resulta de hacer:

$$R(1,1) + R(1,2) + R(1,3) + R(1,4) = 3.775$$

Que se obtiene al reorganizar la matriz y calcular $(I - Q)^{-1}$:

$$P = \begin{matrix} & \begin{matrix} A & R & 1 & 2 & 3 & 4 \end{matrix} \\ \begin{matrix} A \\ R \\ 1 \\ 2 \\ 3 \\ 4 \end{matrix} & \left(\begin{array}{cccccc} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 6/100 & 0 & 3/100 & 91/100 & 0 & 0 \\ 6/100 & 0 & 0 & 3/100 & 91/100 & 0 \\ 4/100 & 0 & 0 & 0 & 3/100 & 93/100 \\ 4/100 & 93/100 & 0 & 0 & 0 & 3/100 \end{array} \right) \end{matrix}$$

Ejercicio 9

Enunciado:

El modelo Wright-Fisher describe la evolución de una población fija de k genes. Los genes pueden ser de uno de dos tipos, llamados alelos: A o a. Sea X_n el número de alelos A en la población en el momento n , donde el tiempo se mide por generaciones. Bajo este modelo, el número de alelos A en el momento $n + 1$ se obtiene muestreando con reemplazo desde la población de genes en el momento n . Por lo tanto, habiendo i alelos de tipo A en el momento n , el número de alelos A en el momento $n + 1$ tiene una distribución binomial con parámetros k y $p = i/k$. Esto resulta en una cadena de Markov con matriz de transición definida por:

$$P_{ij} = \binom{k}{j} \left(\frac{i}{k}\right)^j \left(1 - \frac{i}{k}\right)^{k-j}, \text{ para } 0 \leq i, j \leq k.$$

- (a) Simular este proceso para algún valor de k .
- (b) Observar qué valor toma P_{00} y P_{kk} .
- (c) Cuando la cadena progresa, la población, en algún momento, termina con todos los alelos a (estado 0) o todos los alelos A (estado k). Determinar cuál es la probabilidad de que la población evolucione al estado k .

Resolución:

(a) Tomando $k = 5$:

Código:

```
1 Pij <- function(i,j,k)
2   choose(k,j) * (i/k)^j * (1-(i/k))^(k-j)
3
4 P = matrix(rep( 1, len=36), nrow = 6)
5
6 for(i in 1:6)
7   for(j in 1:6)
8     if(!(i==1 && j==1) && !(i==6 && j==6))
9       P[i,j] = Pij(i-1,j-1,5)
10
11 states <- c("0","1","2","3","4","5")
12 transition <- P
13 mc <- new("markovchain",states = states
14           ,transitionMatrix = transition)
15 mc
16 rmarkovchain(mc,n=10)
```

Resultados:

```
1      0      1      2      3      4      5
2 0 1.00000 0.0000 0.0000 0.0000 0.0000 0.0000
3 1 0.32768 0.4096 0.2048 0.0512 0.0064 0.00032
4 2 0.07776 0.2592 0.3456 0.2304 0.0768 0.01024
5 3 0.01024 0.0768 0.2304 0.3456 0.2592 0.07776
6 4 0.00032 0.0064 0.0512 0.2048 0.4096 0.32768
7 5 0.00000 0.0000 0.0000 0.0000 0.0000 1.00000
8
9 > rmarkovchain(mc,n=10)
10 [1] "2" "2" "3" "2" "1" "0" "0" "0" "0" "0"
```

(b)

$$P_{00} = \binom{k}{0} \left(\frac{0}{k}\right)^0 \left(1 - \frac{0}{k}\right)^{k-0} = 1 * 0^0 * 1^k$$

$$P_{kk} = \binom{k}{k} \left(\frac{k}{k}\right)^k \left(1 - \frac{k}{k}\right)^{k-k} = 1 * 1^k * 0^0$$

Como se puede observar se llega a un resultado el cual no esta definido. Entonces intentemos entender como se calculan las probabilidades.

Si se tiene que $k = 10$ y $X_n = 0$ es decir que la población en la n -ésima generación tiene los genes *aaaaaaaaaa*, para calcular la probabilidad de que $X_{n+1} = 0$ se calcula la probabilidad de formar la palabra *aaaaaaaaaa* tomando letras con reemplazo de la palabra anterior, como la probabilidad de tomar una *A* es 0 ya que lo único que hay son letras *a*, tomando 10 letras con reposición de la palabra *aaaaaaaaaa* la única palabra posible es *aaaaaaaaaa*. Por lo tanto la probabilidad de obtener una palabra sin *A* es 1, es decir que $P_{00} = 1$. Análogamente con *AAAAAAAAAA* se puede llegar a que $P_{kk} = 1$

(c) **Código:**

```

1 simulacion <- function(x0){
2   r = rep(0,1000)
3   for(i in 1:1000){
4     x = rmarkovchain(mc,n=100,t0 = x0)
5     for(j in 1:100){
6       if(x[j] == "5"){
7         r[i] = 1
8         break
9       }
10    }
11  }
12  mean(r)
13 }
```

Resultados:

```

1 > simulacion(4)
2 [1] 0.8
3 > simulacion(3)
4 [1] 0.59
5 > simulacion(2)
6 [1] 0.399
```

Haciendo la simulación de 1000 trazas estimamos la probabilidad de llegar a 5 partiendo de 4, 3 y 2 alelos *A*.

Parte Teórica:

Buscamos $F(i, 5)$ siendo i un estado transitorio, para eso calculamos:

$$G = S * B$$
$$S = (I - Q)^{-1}$$
$$\begin{matrix} & 0 & 5 \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \end{matrix} & \begin{pmatrix} 0.8 & 0.2 \\ 0.6 & 0.4 \\ 0.4 & 0.6 \\ 0.2 & 0.8 \end{pmatrix} \end{matrix}$$

Donde $F(4, 5) = 0.8$, $F(3, 5) = 0.6$, $F(2, 5) = 0.4$

Ejercicio 10

Enunciado:

El día de la elección, las personas participaron en un centro de votaciones de acuerdo con un proceso de Poisson. El promedio, 100 votantes llegan cada hora.

- Si 150 personas arribaron durante la primera hora ¿Qué probabilidad hay de que al menos 350 votantes arriben antes de las 3 horas?
- Simular dicho proceso y graficar una trayectoria.
- Considerar la variable aleatoria: tiempo entre arribos. Obtener los valores a través de simulación y graficar. Determinar qué distribución tiene el tiempo entre arribos de votantes.

Resolución:

- Aquí podemos aplicar 2 de las hipótesis fundamentales de los procesos de Poisson:
 - H_1 : El número de votantes que llegan durante intervalos de tiempo no sobrepuestos son variables aleatorias independientes.
 - H_2 : Si X_t se define como "La cantidad de votantes que llegaron luego de haber transcurrido t horas" y si Y_t es el número de votantes que llegaron entre $[t_1, t_1 + t)$, entonces X_t y Y_t tienen la misma distribución de probabilidad. Es decir, la cantidad de votantes que llegue en un intervalo dado sólo depende de la longitud del intervalo en horas.

$$P(X_3 \geq 350 | X_1 = 150) = P(X_3 - X_1 \geq 200) = 1 - P(X_3 - X_1 < 200) =$$
$$= 1 - (P(Y = 0) + P(Y = 1) + \dots + P(Y = 199))$$

donde $Y \sim Po(2 * 100)$

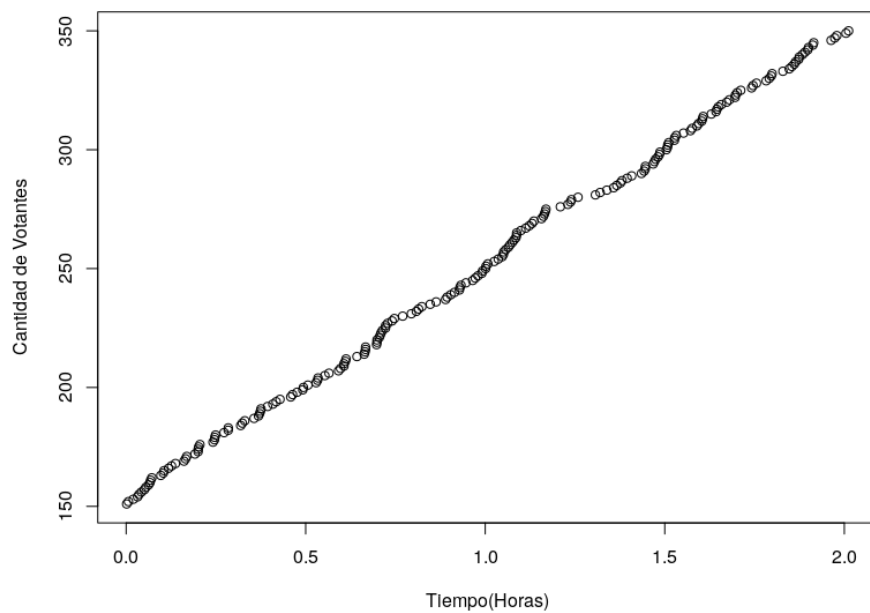
Código:

```
1 r = 0
2 for(i in 0:199)
3   r = r + dpois(i, lambda=200)
4 1-r
```

Resultados:

```
1 [1] 0.5094034
```

- (b) Se simula dicho proceso en el tiempo obteniendo la siguiente gráfica donde se muestra la cantidad total de votantes que llegaron hasta el momento en las 2 horas siguientes:

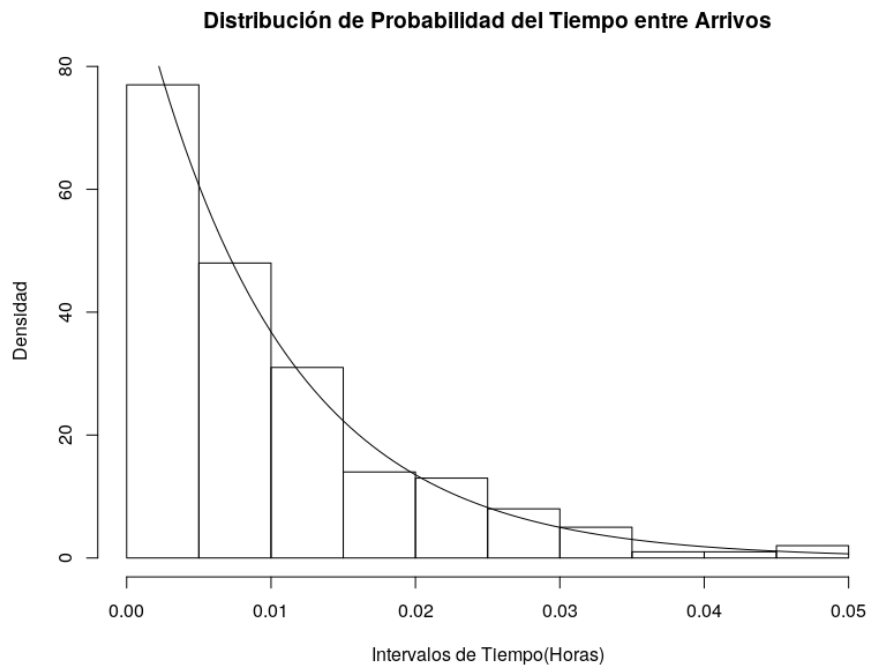


- (c) Luego sacando la diferencia que hay entre los arriros de los votantes en la simulación podemos representar la distribución de probabilidad de la longitud de los intervalos mediante un histograma de densidad.

Código:

```
1 hist(dif,freq = FALSE,  
2      xlab = "Intervalos de Tiempo(Horas)",  
3      ylab = "Densidad",  
4      main = "Distribución de Probabilidad del Tiempo  
5           entre Arriuos")  
5 curve(dexp(x,100),add = TRUE)
```

Resultados:



Conclusión:

Comparando el histograma con la curva de distribución de probabilidad de la distribución exponencial, podemos observar que la longitud de los intervalos entre votantes se distribuye con una distribución exponencial $Exp(100)$ donde la probabilidad de densidad puntual es:

$$f(x) = 100 * e^{-100*x}$$